Capacity design, organizational structure and differential treatment

Emil Temnyalov*

February 12, 2023

Preliminary draft. Please do not circulate.

Abstract

This paper studies the effect of capacities and organizational structure on differential treatment in labor markets and in school choice. Agents have unobservable payoff-relevant types, observable signals and non-payoff-relevant characteristics. A principal assigns the agents within an organizational hierarchy (a firm's hierarchy or a school district) with different tiers and different capacities across tiers, to maximize revenue or efficiency. I first introduce new metrics to quantify how much differential treatment is observed within the organization. I then introduce a taxonomy to compare organizational hierarchies. Finally, I show that flatter, more differentiated, and more "top-heavy" hierarchies induce less differential treatment, under various conditions on the agents' signal distributions. I discuss the implications of the results for the optimal design of firm hierarchies and school districts. These results can also be applied to other vertically differentiated and capacity constrained settings that feature differential treatment.

Keywords: capacity design, organizational structure, optimal assignment, differential treatment, discrimination, school choice, firm hierarchy. JEL: D47, D61, D8, J7, I24.

*Temnyalov: University of Technology Sydney, Department of Economics, PO Box 123, UTS Business School, Sydney, NSW 2007. emil.temnyalov@uts.edu.au.

This is a work in progress and the draft is preliminary.

For helpful comments I thank seminar and conference audiences at BI Oslo, Ecole Polytechnique, Hitotsubashi University, the University of Auckland, University College Dublin, Public Economic Theory 2022, Society for the Advancement of Economic Theory 2022, and the Econometric Society Tokyo 2022 Meeting.

1 Introduction

How does the structure of an organization, for instance a firm or a school district, determine the extent to which we observe differential treatment of the agents within that organization? I consider a setting where the organization observes signals of agents' payoff-relevant types, along with some non-payoff relevant characteristics, and must then assign the agents across different tiers within its organizational structure in order to maximize revenue or efficiency. In education and labor market contexts, such as school choice or hiring and promotions within firm hierarchies, the participants may be treated differently based on their characteristics, e.g. their ethnicity, gender, or other socio-economic status (SES) covariates. That is, an agent's assignment within the organizational structure, as a function of their signal, may vary across characteristics. For example, when students are admitted to selective and non-selective schools within a school district based on their admission test scores, admission criteria may vary across socio-economic status. Similarly, when employees are hired or promoted at different tiers within a firm based on their past productivity, assignments may vary across gender depending on the firm's gender equity policies.

The extent of differential treatment in such settings indirectly depends on the organization's structure. That is, the different tiers of the hierarchy and the number of positions within tiers can systematically affect how much differential treatment one observes in the optimal assignment that is induced by the organization's objectives. Understanding this interaction between organizational structure and assignment policies is worthwhile from both a descriptive and a normative perspective. In the former sense, one would want to better understand in which settings there may be more or less inequality in outcomes across agents' characteristics. In the latter sense, this question can inform the design of organizations so as to reduce disparities and promote fairness.

To study these questions I develop a model based on the general and non-parametric differential treatment framework of Temnyalov (2021). In this setup a principal (e.g. a school district superintendent or a decision-maker in a firm) wants to assign agents (students or employees) to different positions within the organization's hierarchy (e.g. seats in selective and non-selective schools within the district or jobs within the firm's various employment tiers). The principal's objective is to maximize surplus (e.g. the social surplus generated within the school district or the profit of the firm). The surplus that an agent generates in any position depends on their type (a student's ability or an employee's productivity), while the principal only observes a noisy signal of this type (e.g. a student's admission test score or the past output of the employee), as well as some socio-economic or demographic characteristics (e.g. household income, ethnicity or gender). While the latter are not directly surplus-relevant, the mapping between agents' types and signals can vary across characteristics. The principal's optimal policy therefore takes this into account, as the principal forms beliefs about the agents' types based on both their signals and their characteristics.

Temnyalov (2021) characterizes the optimal assignment policy in a setup with multiple organizational tiers and different capacities across tiers, and with no parametric assumptions on the surplus function or on the distributions of agents' types and signals. The optimal assignment is monotone with respect to a particular benefit index, which measures the expected incremental gains from assigning an individual to a higher tier, but it is generally not monotonic with respect to signals or types. This policy features differential treatment because the distributions of types conditional on signals vary across characteristics. Agents may be treated differently not only with respect to their signals, but also in terms of the induced mapping between their unobserved types and their assignments. Hence differential treatment can refer to both statistical discrimination and also to a surplus-maximizing policy that compensates for statistical differences in the distributions of agents' signals.

To understand the interaction between organizational structure and differential treatment, the first challenge is how to quantify the latter. In this paper I introduce three metrics which measure the extent of differential treatment in an organization from the perspective of an observer who sees the agents' signals and characteristics and their assignments. The first metric compares across characteristics the signals of the marginal agents assigned to each tier. The second metric compares across characteristics the signals of the representative agents assigned to each tier. Finally, the third metric compares across characteristics the summary statistics of the signals of all agents assigned to each tier.

Next, I introduce a taxonomy of organizational structures, which classifies them in terms of how flat or how hierarchical they are. As special cases, this taxonomy also captures the notions of "top-heavy" and "bottom-heavy" hierarchies. It enables me to compare the three measures of differential treatment across organizations, in terms of how differentiated positions are across the tiers of the organization.

I show that flatter organizations will generally display less differential treatment according to all three of my metrics, under a condition on the agents' signal distributions called monotone signal differences. This condition means that agents from one category have increasingly higher signals relative to agents from another category, conditional on having the same benefit index. This condition is plausible in many contexts: for example, it can be satisfied when one group of agents is over-represented at higher signals, compared to another group of agents. The result means that for any organizational tier at which signal differences are positive, re-allocating positions upwards reduces the extent of differential treatment. For any tier at which signal differences are negative, re-allocating positions downwards reduces the extent of differential treatment. That is, there exists some intermediate reference tier, such that flattening the organization relative to that tier induces less differential treatment. In the context of a firm's tiered employment structure, for example, this result suggests that de-layering the firm's hierarchy can reduce the gender gap.

I also show that if the agents' categories or characteristics can be ordered, based on first-order stochastic dominance of the conditional type distributions, then more "top-heavy" organizational hierarchies induce less differential treatment. Specifically, the ordering over characteristics requires that one group of agents has better type distributions (in the sense of FOSD) than another group, conditional on them having the same signals. Such an assumption can realistically apply to, for example, household income in a school admission setting. That is, students with higher household income may have better access to test preparation, and as a result, conditional on achieving the same admission test score, a higher-income student may have a worse type distribution, compared to a lower-income student. In such a setting, re-allocating positions upwards, i.e. making the hierarchy more top-heavy, always reduces all three of my metrics of differential treatment. In the context of school admissions, this result suggests that expanding the capacities of selective schools can systematically reduce the inequality of access across socio-economic status.

Finally, these results imply that organizational *design* can be a useful policy tool to reduce the extent of economic inequality. Organizations in many contexts have tried to reduce disparities across agents directly, by adopting different assignment policies. For example, a school district might adopt quotas and reserves for students of different socio-economic categories or an affirmative action policy for minority students, in order to address the under-representation of some groups of students and reduce inequality in outcomes. Similarly, a firm might adopt gender-based hiring and promotion policies in order to reduce disparities between men's and women's positions in the organization. In such contexts, understanding what organizational structures yield more or less differential treatment can enable the organization to design its hierarchy in a way that inherently reduces the inequality that is observed across socio-economic or demographic characteristics, above and beyond the impact of any preferential treatment or affirmative action policies.

2 Related literature

The main focus of this paper is to study how organizational structure affects differential treatment. While previous work has studied different forms of differential treatment, such as controlled school choice and statistical discrimination, these literatures generally assume a fixed and stylized organizational structure. This is the first paper to study how differential treatment varies across different organizational structures, from an economic perspective. I explore this in a general and non-parametric theoretical framework, which allows for a wide range of type and signal distributions, as well as different organizational hierarchies.

In the market design literature, papers on controlled school choice and other forms of differential treatment (Abdulkadiroğlu and Sŏnmez, 2003; Abdulkadiroğlu, 2005; Hafalir, Yenmez, and Yildirim, 2013; Ehlers, Hafalir, Yenmez, and Yildirim, 2014) assume that the set of schools and positions that students are matched to is an exogenous input into the matching process. That is, the set of schools and their associated capacities or seats, which constitute the organizational structure of a school district, are fixed. A natural next step is to ask how the extent of differential treatment that is observed in a controlled school choice setting would vary if the structure of the school district changed. Here, I consider this type of question in an assignment game. In my model differential treatment arises endogenously, rather than through fixed quotas or reserves. One can think of such quotas and reserves as a particular implementation of some optimal assignment policy, but in my framework I can more directly study the extent of differential treatment as a function of the set of schools and positions. My main question here is how the optimal assignment of agents to positions varies across different organizational structures, and in particular how the structure affects the extent of differential treatment induced by the assignment policy.

Dur and Van der Linden (2021) are the first to study capacity design as a policy lever in school choice, as opposed to an external input. The focus of this work however is different: it studies how to re-allocate seats across schools based on students' preferences, in a way that is efficient and respects students' priorities, whereas I consider how the structure of capacities affects differential treatment.

A smaller subset of the school choice literature takes a similar approach to school district *priorities* as I do here in terms of school district *structure*. Rather than considering priorities as arbitrary inputs into the matching process, Ergin (2002) and Ehlers and Erdil (2010), among others, ask what conditions on the priorities yield some desirable properties such as efficiency and strategy-proofness; Dur, Kominers, Pathak, and Sönmez (2018) study the design of the precedence order in which reserves are processed; and Celebi and Flynn (2020) study the optimal design of coarse priorities that are derived from some finer input. Here, I consider the school district's structure in a similar way: I evaluate different possible structures according to the extent of differential treatment they induce, and propose that the structure itself can be viewed as a design instrument to reduce inequality.

This paper also relates to the literature on statistical differentiation, which is a form of differential treatment. Starting with the seminal works of Phelps (1972) and Arrow (1973), two strands of papers have explored how discrimination can arise in contexts such as education or labor markets, based on exogenous or endogenous differences between groups of agents. These include, among many others, Aigner and Cain (1977); Lundberg and Startz (1983); Coate and Loury (1993); Mailath, Samuelson, and Shaked (2000); Moro and Norman (2004)[1] This literature generally considers a very stylized organizational structure, where for example a firm either hires or does not hire a worker, as the main focus is on explaining discrimination, rather than the effect of the firm's hierarchy on discrimination. A small subset of papers consider statistical discrimination in hierarchical settings: for example, Fryer (2007) considers a two-tier firm and studies statistical discrimination in both hiring and promotions; Bjerk (2008) considers a model with 3 tiers of jobs and also studies discrimination in promotions. While these papers model a firm hierarchy that consists of more than a single tier of workers, they do not study how discrimination varies across different hierarchies, e.g. across hierarchies with different job tiers and different numbers of positions within tiers. In contrast, here I introduce a taxonomy of organizational structures and compare the extent of differential treatment across structures. This also suggests a role for deliberate organizational design as a policy instrument to reduce discrimination.

Finally, this paper relates to several prominent findings in the management and economics literatures on organizational structure and design. Previous empirical work has documented a pattern of de-layering of firm hierarchies over time (Rajan and Wulf, 2006; Bloom and Van Reenen, 2007; Colombo and Delmastro, 2008; Wulf, 2012, among others). A small theoretical literature has sought to explain these developments: e.g., Rajan and Zingales (2001) develop a model where the organizational hierarchy bifurcates into top and bottom tiers, at the expense of a middle management tier. This is the first paper to consider the effect of such organizational re-structuring on differential treatment, in an informational economics framework where the principal may discriminate among agents based on non-payoff relevant characteristics.

[1] Fang and Moro (2011) provide a comprehensive survey.

3 Model

A principal (e.g. a decision-maker in a firm or a school district superintendent) wants to assign agents (e.g. employees or students) to positions within an organizational structure. There is a set of agents N, with size $|N| = n$. The organization's structure consists of multiple tiers (e.g. job levels or different types of vertically differentiated schools, such as selective and non-selective schools), denoted by $P = (p_1, ..., p_m)$, with corresponding capacities $(k_1, ..., k_m)$, which represent the number of positions within each tier. Without loss of generality I assume that $\sum_j k_j \geq n$, as one can think of p_1 as a null tier with unlimited capacity, which represents not being assigned to any tier of the organization.

Each agent $i \in N$ has an unknown type $t_i \in T$, drawn from a distribution F, where the type space T is an ordered set with discrete or continuous values. The principal observes each agent's characteristics, or the categories the agent belongs to, denoted $x_i \in X$, where the characteristics space X can be single- or multi-dimensional, can take discrete or continuous values, and need not be ordered. The principal also observes a signal of the agent's type, $s_i \in S$, drawn from a distribution F_{t_i,x_i}, where the signal space S is an ordered convex set. Here the agent's type can represent the academic ability of a student or the productivity of an employee, and the signal can represent the student's admission test score or the employee's past output. Both the type and the signal therefore are naturally ordered. On the other hand, the agent's characteristics can represent both socio-economic status covariates (e.g. a student's household income) and demographic covariates (e.g. a student's ethnicity or an employee's gender), so they need not be ordered.

An individual agent generates some ex post surplus that depends on their type and their position, denoted by $v : T \times P \to \mathbb{R}$. This surplus can refer to either social surplus or profit, depending on the setting. In the education context, it represents the social surplus generated by admitting a student to a particular school; in the firm context, it represents the profit that the employee generates when employed at a particular tier of the firm's hierarchy. I assume that v is increasing in both arguments and supermodular, i.e. $v(t'', p'') + v(t', p') > v(t'', p') + v(t', p'')$ for all $t'' > t'$ and $p'' > p'$. The latter means that the agent's type and their position are complements. This is a natural assumption in the education and labor market contexts, and is also the usual assumption made in the literature. It ensures that the first-best assignment, if types are observable, features positive assortative matching, i.e. higher-type agents are optimally assigned to higher tiers.

The principal's policy is an assignment that maps the agents' observable signals and characteristics to organizational tiers, denoted by $\mathscr{P} : (X \times S)^n \to \Delta(P^n)$. I denote by $\mathbb{P}(p|x_i, s_i)$ the distribution of agent i's assignment and by $p(x_i, s_i)$ the agent's realized assignment. Furthermore, the signal distributions $\{F_{t,x}\}$ and the assignment policy $\mathscr{P}$ jointly induce some indirect assignment mapping $\mathbb{A} : T \times X \to \Delta(P)$, where $\mathbb{A}(t_i, x_i)$ denotes the distribution of agent i's position as a function of their type and characteristics.[2]

The principal's objective is to design a policy that maximizes the expected total surplus from the assignment, subject to ex post feasibility:

$$\begin{aligned} \max_{\mathscr{P}} \quad \sum_i \mathbb{E}[v(t_i, p(x_i, s_i))] \quad &= \quad \sum_i \sum_j \int_T v(t_i, p_j) dF(t_i|x_i, s_i) \cdot \mathbb{P}(p_j|x_i, s_i) \\ \text{s.t.} \quad |\{i : p(x_i, s_i) = p_j\}| &\leq k_j \, \forall j \end{aligned} \tag{1}$$

In the education context, for example, this objective function means that the school district wants to maximize the aggregate returns to education. In the labor market context, it means that the firm seeks to maximize total profit.

3.1 Assumptions

To build onto the results in Temnyalov (2021), I make some assumption on the type and signal distributions within and across characteristics. First, I assume the expected surplus

$$V_{p,x}(s) := \mathbb{E}[v(t,p)|x, s]$$

is continuous in s. This is a mild technical assumption, because it is only imposed conditional on x. That is, the agent's expected surplus in position p varies smoothly as a function of their signal s, holding x constant.[3]

Second, I assume that the family of signal distributions $\{F_{t,x}\}$ satisfies the strict monotone likelihood ratio property (MLRP) in t, for fixed x. I.e. for $s'' > s'$, the ratio of the densities $\frac{f_{t,x}(s'')}{f_{t,x}(s')}$ is increasing in t, conditional on x, whenever it is well-defined. This assumption is standard in the information economics literature and is natural in the education and labor market contexts. It ensures that an agent with a higher signal has a higher type

[2] To simplify the notation I omit the dependence of $\mathbb{P}(\cdot)$, $p(\cdot)$ and $\mathbb{A}(\cdot)$ on $\{(x_j, s_j)\}_{j \in N \setminus i}$, except when necessary for clarity.

[3] Because $F_{t,x}$ can vary across x, two arbitrarily similar signals s and s' for agents with different characteristics x and x' can imply discontinuously different expected surpluses; however, here the continuity assumption is only imposed for agents with the same characteristics, so it is much less demanding.

distribution, in the sense of first-order stochastic dominance (FOSD). Specifically, following Milgrom (1981), this assumption implies that $F(t|x, s'') \succ_{FOSD} F(t|x, s')$ when $s'' > s'$.[4]

Third, I assume characteristics are comparable, as in Temnyalov (2021). Specifically:

Definition 1. *Categories x and x' are **comparable** if for any $s \in S$ one of the following cases holds:*

(i) $\exists s' \in S$ *s.t.* $V_{p,x}(s) - V_{p',x}(s) = V_{p,x'}(s') - V_{p',x'}(s') \, \forall p, p'$;
(ii) $\forall s' \in S$, $V_{p,x}(s) - V_{p',x}(s) \geq V_{p,x'}(s') - V_{p',x'}(s') \, \forall p$ *and* $p' < p$;
(iii) $\forall s' \in S$, $V_{p,x}(s) - V_{p',x}(s) \leq V_{p,x'}(s') - V_{p',x'}(s') \, \forall p$ *and* $p' < p$.

While the previous assumptions hold within characteristics, comparability is an assumption on how signals relate across characteristics. It requires that signals can be compared for agents with different characteristics in terms of the expected surplus that they correspond to, and these comparisons are uniform across different tiers of the organization. Here, agents are compared with respect to their *expected incremental gains* from being assigned to a tier p rather than another tier p', i.e. $V_{p,x}(s) - V_{p',x}(s) = \mathbb{E}[v(t,p)|x,s] - \mathbb{E}[v(t,p')|x,s]$.

Specifically, the first case above means that an agent with characteristics x and signal s has the same expected incremental surplus from being assigned to p rather than p' as an agent with characteristics x' who has some (possibly different) signal s'. In this sense the two agents with (x, s) and (x', s') are equivalent, because they have the same expected incremental surplus, regardless of which two tiers p, p' one uses to carry out the comparison. The second case means that an agent with characteristics x and signal s has larger expected incremental surplus than an agent with characteristics x' and any possible signal s'. This again holds regardless of which tiers p, p' one uses to carry out the comparison. Analogously, the third case means that an agent with (x, s) has smaller expected incremental surplus than an agent with (x', s'), across all p, p' that one can carry out the comparison with. Taken all together, the three cases mean that (x, s) can be compared to any (x', s') in a way that is uniform across any two tiers p and p'.[5]

Comparability is satisfied in many commonly studied settings. For example, Temnyalov (2021) shows that it holds when the organizational tiers are binary, i.e. $P = \{p_1, p_2\}$ (for

[4]This assumption is also only imposed within characteristics, i.e. conditional on a fixed x, which makes it much less demanding than if it were imposed across characteristics. Because $F_{t,x}$ can vary across x, this allows for some $F(t|x', s')$ to first-order stochastically dominate $F(t|x'', s'')$ even if $s' < s''$, when $x' \neq x''$.

[5]Further details and interpretations of comparability are provided in Temnyalov (2021), including two examples described in Figure 1 and in Example 1 in that paper.

instance, when a school district chooses to assign students to selective and non-selective schools); or when the surplus function $v(t, p)$ is multiplicatively separable, as in many mechanism and market design models; or when the signal distributions $\{F_{t,x}\}$ are pairwise equivalent in a specific sense (Temnyalov, 2021, Definition 3).

3.2 Preliminaries

Before considering how differences across characteristics are reflected in the optimal assignment policy, I first formally define and distinguish differential treatment with respect to signals and types.

Definition 2. *Consider two assignment distributions* $\mathbb{P}(p|(x_i = x, s_i = s), \{(x_j, s_j)\}_{j \in N \setminus i})$ *and* $\mathbb{P}(p|(x_i = x', s_i = s), \{(x_j, s_j)\}_{j \in N \setminus i})$, *which differ only based on agent i's characteristics.*

Categories x and x' are ***treated differently with respect to signals*** *if for some $s \in S$ and* $\{(x_j, s_j)\}_{j \in N \setminus i}$, $\mathbb{P}(p|(x_i = x, s_i = s), \{(x_j, s_j)\}_{j \in N \setminus i}) \neq \mathbb{P}(p|(x_i = x', s_i = s), \{(x_j, s_j)\}_{j \in N \setminus i})$.

Categories x and x' are ***treated differently with respect to types*** *if for some $t \in T$ and* $\{(x_j, s_j)\}_{j \in N \setminus i}$, $\mathbb{A}(t_i = t, x_i = x|\{(x_j, s_j)\}_{j \in N \setminus i}) \neq \mathbb{A}(t_i = t, x_i = x'|\{(x_j, s_j)\}_{j \in N \setminus i})$.

The optimal policy which maximizes the organization's value is given by the following result, adapted from Temnyalov (2021).[6]

Proposition 1. *Under the assumptions above, the optimal policy $\mathscr{P}^*$ assigns agents to tiers assortatively with respect to a* benefit index *defined by* $r(x, s) := \mathbb{E}[v(t, p_m)|x, s] - \mathbb{E}[v(t, p_{m-1})|x, s]$.

This optimal assignment policy features differential treatment for two distinct reasons. On one hand, signals may be biased for some categories of agents. For example, a signal s for an agent with characteristics x may correspond to the same benefit index as a signal s' for an agent with different characteristics x'. When $s > s'$ the signal is biased upwards for category x, while when $s < s'$ the signal is biased downwards for category x. In such cases the optimal policy features differential treatment with respect to the agents' signals. The benefit index corrects for the signal differences across categories, because the comparison of expected incremental gains for agents with different characteristics can directly account for

[6] The optimal assignment rule is deterministic, except for special cases that require tie-breaking among surplus-equivalent agents. To simplicity the exposition here I will assume that there are no such ties, as these do not affect the analysis.

biases. In particular, if signals are biased downwards for a category x relative to x', then category x is treated more favorably by the optimal policy, because the benefit index is larger for x, conditional on the same signal s for category x'.

On the other hand, differential treatment may also arise because signals are differentially informative. For example, suppose category x has noisier signals than category x', in the sense that $F(t|x, s)$ is a mean-preserving spread of $F(t|x', s)$ for all s. Under some conditions on the surplus function $v(t, p)$, the benefit indices are not equal across categories, i.e. $r(x, s) \neq r(x', s)$, even though $\mathbb{E}[t|x, s] = \mathbb{E}[t|x', s]$.[7] For example, if $r(x, s) < r(x', s)$, then category x' is treated more favorably by the optimal policy, because its signals are more informative about the agents' types. Hence differential treatment is again observed with respect to signals, since $r(x, s) \neq r(x', s)$. But in addition to this, the policy may also feature differential treatment with respect to the agents' unobserved types. In particular, with differentially informative signals the indirect assignment mapping $\mathbb{A}$ is generally not constant in x, i.e. $\mathbb{A}(t, x) \neq \mathbb{A}(t, x')$, even though signals are unbiased, and hence the optimal policy leads to aggregate inequality among the different categories of agents.[8]

The two sources of differential treatment described above, relating to bias and to differential informativeness, can co-exist in some contexts, and may even induce differential treatments with respect to signals and with respect to types which go in opposite directions. The following example illustrates this situation in a school admission application.

Example 1. *Suppose a school district needs to assign two students to two seats in a selective and a non-selective school; i.e. $P = (p_1, p_2)$, where p_1 represents the non-selective school and p_2 the selective school, each with 1 position, $k_1 = k_2 = 1$. The students $i \in \{1, 2\}$ have normally distributed abilities, $t_i \sim N(0, 1)$, and have different characteristics, x_1 and x_2 respectively. A student of category $j \in \{1, 2\}$ has an admission test score $s_i = t_i + \varepsilon_i^j$. Suppose $\varepsilon_i^1 \sim N(-\delta, \sigma_1^2)$ and $\varepsilon_i^2 \sim N(0, \sigma_2^2)$. That is, category x_1's signal is biased downwards by some $\delta > 0$, relative to category x_2, and the signals are differentially informative if $\sigma_1^2 \neq \sigma_2^2$.*

First, suppose $\sigma_1^2 = \sigma_2^2$. In this case the benefit indices are straight-forward horizontal translations across categories, i.e. $r(x_1, s) = r(x_2, s + \delta)\ \forall s$, since $F(t|x_1, s) = F(t|x_2, s + \delta)\ \forall s$. The optimal policy treats the two agents differently with respect to their signals, because it fully off-sets the signal bias and category x_1 is favored by an amount that reflects δ.

Next, suppose $\sigma_1^2 > \sigma_2^2$, so category x_1's signals are biased downwards and also noisier

[7] Specifically, if v has either strictly convex or strictly concave differences in t (Temnyalov, 2021, Definition 4), then $r(x, s) > r(x', s)\ \forall s$ or $r(x, s) < r(x', s)\ \forall s$, respectively.

[8] This is illustrated in a specific parametric setting in Temnyalov (2021), Example 5.

than those of category x_2. If $v(t,p_2) - v(t,p_1)$ is a strictly concave function of t, then $r(x_1,s) < r(x_2,s+\delta)$ $\forall s$, since $F(t|x_1,s)$ is a mean-preserving spread of $F(t|x_2,s+\delta)$ $\forall s$. Note also that $F(t|x_2,s+\delta) \succ_{FOSD} F(t|x_2,s)$ $\forall s$, so $r(x_2,s+\delta) > r(x_2,s)$ $\forall s$. Therefore it is not immediately clear whether the optimal policy treats category x_1 more favorably with respect to signals than category x_2. This depends on the comparison between $r(x_1,s)$ and $r(x_2,s)$ for different values of s, which can go in either direction, depending on the magnitudes of δ, σ_1^2 and σ_2^2. There exists some threshold $\bar{s}_2(s_1+\delta)$, which is a function of s_1 and δ, with the properties that $\bar{s}_2(s_1+\delta) < s_1+\delta$ and $r(x_1,s_1) = r(x_2,\bar{s}_2(s_1+\delta))$. Category x_1 is favored with respect to signals if $\bar{s}_2(s_1+\delta) > s_1$, and conversely category x_2 is favored with respect to signals when $\bar{s}_2(s_1+\delta) < s_1$. In particular, it is possible that category x_2 is treated more favorably with respect to signals, relative to category x_1, even though the latter has signals that are biased downwards.

Regardless of which category is favored with respect to signals, in this case one can unambiguously show that category x_2 is treated more favorably with respect to types, and in the aggregate. Consider the induced indirect assignment mappings $\mathbb{A}(t,x_1)$ and $\mathbb{A}(t,x_2)$. Conditional on a type t, the probability that student 2 is admitted to p_2 is $P(s_2 \geq \bar{s}_2(s_1+\delta)) = \int P(s_2-(s_1+\delta) \geq \bar{s}_2(s_1+\delta)-(s_1+\delta))dF(s_1)$. Notice that $s_2-(s_1+\delta)$ is normally distributed and has mean 0, while $\bar{s}_2(s_1+\delta)-(s_1+\delta) < 0$. Therefore the integrand is larger than $\frac{1}{2}$ everywhere and $P(s_2 \geq \bar{s}_2(s_1+\delta)) > \frac{1}{2}$. Thus in aggregate the optimal assignment rule treats the two categories of agents differently with respect to their types. Analogously, this observation also holds conditional on $t_1 = t_2$, i.e. a category x_2 student with type t is strictly more likely to be admitted than a category x_1 student of the same type type t. Hence category x_2 is favored with respect to types, both in the aggregate and when comparing same-type students.

Finally, I define some additional notation which will be useful later. Consider an organizational structure $(P;K) = (p_1,...,p_m;k_1,...,k_m)$, and the assignment induced by some policy $\mathscr{P}$. For any position p_j, denote the set of agents who are assigned to p_j under $\mathscr{P}$ by

$$I_j := \{i : p(x_i,s_i) = p_j\}.$$

The corresponding signals and characteristics of these agents are $\{(x_i,s_i)\}_{i\in I_j}$. Denote the set of benefit indices of these agents by

$$R_j := \{r(x_i,s_i) : i \in I_j\}.$$

Lastly, define the inverse of $r(x,s)$ as follows: $s(x,r) := \{s : r(x,s) = r\}$.

4 Analysis

Given the characterization of the optimal policy above, I now examine how the organization's structure determines the extent of differential treatment. To study this, I first define three metrics which quantify differential treatment and discuss their interpretations. I then introduce a taxonomy of organizational structures, which allows me to compare organizations depending on their features or properties, such as the number of tiers in the hierarchy and the distribution of positions across tiers.

4.1 Measures of differential treatment

My first metric of differential treatment starts with the following question: what signal would an agent with characteristics x need to have, in order to be marginally assigned to tier p_j under the optimal assignment policy $\mathscr{P}^*$? To quantify the extent of differential treatment between any two categories x and x', at any tier p_j, one can compare these hypothetical minimal signals corresponding to x and x'. Whenever these signals associated with a tier p_j differ across x and x', this represents differential treatment. One can then aggregate the tier-specific differences across all the tiers of the organization, into an aggregate measure of differential treatment between x and x', by summing up the absolute values of differences in treatments at each tier. This yields the following metric.

Definition 3. *Consider categories x and x'. The* ***first metric of differential treatment*** *is*

$$DT_1(x, x') := \sum_j |s(x, \min\{R_j\}) - s(x', \min\{R_j\})|$$

To elaborate on this, consider a setting where the agents have some specific signals and characteristics, $\{(x_i, s_i)\}_{i \in N}$. Suppose that some agent i is the one who is marginally assigned to some particular tier p_j, under the organization's optimal policy $\mathscr{P}^*$. That is, i has the lowest benefit index $r(x_i, s_i)$ among the agents assigned to p_j. The question is, how high of a signal would an agent with characteristics x need to have, in order to displace agent i from tier p_j? Similarly, how high of a signal would an agent with characteristics x' need to have, in order to displace i? The difference between these, aggregated in absolute values across all tiers, is precisely the definition of $DT_1(x, x')$.

This metric has a natural interpretation in the school choice and firm promotion contexts, for example. In a firm where employees are promoted based on their past output, one could ask how much output would a female employee need to have in order to be marginally promoted to the next tier of the firm's hierarchy. If the answer is different than that for a male employee, then the difference provides a way to quantify the extent of differential treatment.[9] In a school district which assigns students to selective and non-selective schools based on merit rankings and diversity policies, one could ask how high of a merit score a low-SES student would need to have in order to be marginally admitted to a selective school, and compare that to how high of a score a high-SES student would need.[10]

Next, my second metric of differential treatment starts with the question: what signal would a representative agent in tier p_j have, if they have characteristics x? To quantify the extent of differential treatment between any two categories x and x', at any tier p_j, one can compare these hypothetical signals of the representative agent across x and x'. As with the previous metric, these differences across tiers can then be aggregated in absolute values across all tiers, yielding the next metric.

Definition 4. *Consider categories x and x'. The* ***second metric of differential treatment*** *is*

$$DT_2(x, x') := \sum_j |s(x, \bar{R}_j) - s(x', \bar{R}_j)|,$$

where $\bar{R}_j$ is any summary statistic of R_j.

As before, suppose the agents have some specific signals and characteristics, $\{(x_i, s_i)\}_{i \in N}$. Consider the "typical" agent in any tier p_j—typical in the sense that this agent has a benefit index that is representative for all the agents assigned to tier p_j. This may for instance be the average benefit index or the median benefit index of agents assigned to p_j. What signal would this agent have if they have characteristics x or x'? The aggregated difference between these yields the definition of $DT_2(x, x')$.

[9] In this context, an employee's past output may be systematically affected by parental leave to different extents across genders: if women shoulder more of the burden of parenting, then their careers face more significant disruptions. The mapping between female employees' types (their productivity or ability) and their signals (their past output) would therefore be different than those of male employees. The firm's profit-maximizing promotion or hiring policy would account for these differences through the employees' benefit indices.

[10] In this context, a student's benefit index reflects both their admission test score and their other characteristics. Admission policies that account for other characteristics, such as socio-economic status or demographic characteristics, are assumed to capture the differences in benefit indices across different categories of students. That is, the school district can treat students with (x, s) and (x', s) differently, because their different characteristics affect the mapping between types (e.g. their returns to education) and signals (their admission test score). Access to costly after-school tutoring, for instance, would mean that test scores are biased in favor of high-SES students, so an optimal admission policy should account for this.

In a school choice context, this metric asks what admission score the typical student who is admitted to a selective school would have, if they were a low-SES student or a high-SES student. Here, the representative student at a selective school may be defined as one who has the average returns to a selective school education (relative to a non-selective school education). The comparison then examines what signals these average returns to a selective school education correspond to for agents with different SES characteristics. In a firm context, this metric compares the signals of representative employees of different genders at each tier in the firm's hierarchy, for example. I.e. what would the signal be for the typical employee if they were male or female?

Finally, my third metric compares the representative signals of agents across categories at each tier, as opposed to the signals of marginal or representative agents. Specifically, it asks: what would be the typical signal of the agents in tier p_j, if they had characteristics x or x'? The differences in signals can be measured in terms of a summary statistic which aggregates the signals corresponding to all agents at each tier, taking into account the effect of their characteristics. The differences can then be aggregated across all tiers, yielding the following metric.

Definition 5. *Consider categories x and x'. The **third metric of differential treatment** is*

$$DT_3(x, x') := \sum_j |\bar{S}(x, R_j) - \bar{S}(x', R_j)|,$$

where $\bar{S}(x, R_j)$ is any summary statistic of $\{s(x, r) : r \in R_j\}$.

Unlike the previous two metrics, this one aggregates the signals that correspond to all agents at each organizational tier, taking into account their characteristics, and then compares these aggregated summary statistics across characteristics. That is, this metric considers what signals the agents at tier p_j would have if they had characteristics x or x', i.e. $\{s(x, r) : r \in R_j\}$, and then summarizes these into a statistic $\bar{S}(x, R_j)$, which might for instance be the average of $\{s(x, r) : r \in R_j\}$. The differences in summary statistics for x and x' is then aggregated in absolute values across all tiers of the organization.

In the school choice context, this metric compares what signals the students assigned to a selective school would have, as a function of their characteristics, in terms of average admission test scores for example. This allows one to compare how the typical admission scores would differ if the students admitted to selective schools had low-SES or high-SES characteristics. In the firm context, the metric compares the typical (e.g. average) output (or sales, profit, citations, etc.) of all employees in a tier, if they were female or male, for example.

4.2 Taxonomy of organizational structures

There are many ways in which one can compare organizational structures, in order to study the interplay with differential treatment. In this section I define a taxonomy that is the first step in this direction. Specifically, I focus on the impact of *organizational flatness* on differential treatment. Organizational flattening is a widely documented phenomenon in the economics and management literatures: it refers to the reduction of layers in a firm's hierarchy (Wulf, 2012). Rajan and Wulf (2006); Colombo and Delmastro (2008); Bloom and Van Reenen (2007), among others, document a significant recent pattern in firm structure whereby hierarchies tend to have fewer employee tiers, and study the implications for firm performance.[11] The idea of de-layering is also relevant in the context of education: in school choice, one can compare different school district structures in terms of whether they feature vertically differentiated tiers of schools. Flatter districts are ones with fewer or no vertically differentiated tiers of schools, while more hierarchical districts are ones with more tiers, for example based on selectivity.

In addition to considering the elimination of tiers in the organizational structure, my notion of flatness also allows me to study the re-distribution of positions across tiers, for example the reduction in size of some tiers. In the context of firm hierarchies the number of positions in each tier reflects how "top-heavy" or how "bottom-heavy" a firm is, for example. Guadalupe, Li, and Wulf (2014) find that the number of managers who report directly to the CEO in large US firms has doubled since the 1980s. Rajan and Zingales (2001) develop a theoretical model where the middle tier may shrink and firm hierarchy may bifurcate into top and bottom tiers, consistent with the observation that middle management has generally shrunk in recent decades. In the school choice context, the numbers of positions across ver-

[11] This flattening is generally driven by factors that are not explicitly modeled in this paper, as my focus is on the effect of organizational structures on differential treatment, rather than on the determinants of organizational structure. For example, improvements in monitoring technology, broadly defined, might make it easier for managers to supervise workers directly, rather than delegating this task to a supervisor. Similarly, information technology and automation might altogether make some roles or tiers in the firm obsolete. I consider organizational structure as a primitive of the model and study its implications for differential treatment. My subsequent analysis implies that differential treatment may itself become one of the forces that shape organizations: e.g. gender equity policies in hiring and promotions might affect how positions are distributed across the firm hierarchy, when managers and leaders want to mitigate inequalities across genders.

tically differentiated tiers reflect the capacities of schools with different selectivity. Districts where top-ranked schools have larger capacities are more top-heavy, analogously to the firm hierarchy interpretation.

I define flatter organizations as ones where positions are relatively more concentrated above and below some reference tier of the hierarchy (such as a firm hierarchy that bifurcates into a top and a bottom tier, at the expense of middle management), whereas more hierarchical organizations are those where positions are relatively more concentrated towards the reference tier. My notion of flatness allows me to study both the elimination of tiers (e.g. when a role disappears and positions are subsumed into other tiers) and the re-allocation of positions across tiers (e.g. the shrinking of a middle management tier). It also represents comparisons where one organizational structure is more top-heavy or more bottom-heavy than another. The question is then how such differences across organizational structures affect the extent of differential treatment: for instance, in which organizations would one expect to see more or less gender differences in promotions, and when would one expect to see more socio-economic segregation within selective schools in a school district?

Definition 6. *Consider two organizational structures,* $(P; K)$ *and* $(P'; K')$*, of equal size.* $(P; K)$ *is* ***flatter*** *than* $(P'; K')$ *around the* M^{th} *tier if for some* $M \in \{1, ..., m\}$*,*

$$\sum_{j \leq \widetilde{m}} k_j \geq \sum_{j \leq \widetilde{m}} k'_j \text{ for all } \widetilde{m} < M \text{ and } \sum_{j \geq \widetilde{m}} k_j \geq \sum_{j \geq \widetilde{m}} k'_j \text{ for all } \widetilde{m} > M,$$

and at least one of these inequalities holds strictly for some $\widetilde{m}$*. Conversely,* $(P; K)$ *is* ***more hierarchical*** *than* $(P'; K')$ *around the* M^{th} *tier if the opposite inequalities hold.*

If the above hold for $M = 1$*, i.e. the reference tier* M *is the lowest one, then* $(P; K)$ *is more* ***top-heavy****, and similarly if* $M = m$*, then* $(P; K)$ *is more* ***bottom-heavy****.*

This definition of flatness compares cumulative capacities relative to some reference tier M. When this tier is intermediate, flatness means that positions are more concentrated into "heavier" top and bottom tiers. For instance, an employment structure consisting of a Director, a Manager and a Worker is more hierarchical than a structure with a Director and two Workers.[12] Similar comparisons can also be made when a tier is reduced, rather than eliminated: e.g. an employment structure with a Director, three Managers and three

[12] Throughout the analysis I will disregard differences across organizational structures that are solely due to null-sized tiers. Such differences arise only as a result of different possible labeling of tiers. For example, although $(P = (p_1, p_2, p_3); K = (2, 0, 1))$ and $(P = (p_1, p_2, p_3); K' = (2, 1, 0))$ are technically distinct, I will treat them as identical, since both de facto consist of 2 tiers with equal corresponding capacities in each.

Workers is flatter than a structure with two Directors, a Manager and four Workers. The definition also allows the reference tier be the lowest or the highest in the organization, in which case a flatter organization is one where higher and lower tiers are heavier, respectively.

4.3 Organizational structure and differential treatment

The following results leverage the three metrics of differential treatment and the taxonomy of organizational structures defined above, to consider the effect on differential treatment under different conditions on the agents' signals.

Proposition 2. *Consider two organizational structures $(P;K)$ and $(P';K')$. Suppose some categories x and x' have monotone signal differences, i.e. $s(x,r)-s(x',r)$ is monotone in r.*

There exists some M s.t. if $(P;K)$ is flatter than $(P';K')$ around the M^{th} tier, then $(P;K)$ displays less differential treatment than $(P';K')$ according to all metrics, DT_1, DT_2 and DT_3.

Proposition 2 means that flatter organizational hierarchies will generally display less differential treatment, when some groups of agents have monotone signal differences. From an organizational design perspective, it implies that in such settings one can flatten the organizational structure in a way that will reduce disparities among different groups of agents. In the context of firm hierarchies, flatter hierarchies are ones where positions are more concentrated towards the top and bottom of the hierarchy, away from some intermediate reference tier. This is consistent with the shrinking of the middle management tier that has been documented over several decades in the management literature, for example. It is unlikely that this kind of de-layering of organizations is driven by a push to reduce differential treatment within organizations. Rather, the management literature points to numerous other factors that are likely to have driven these changes. However, the result does mean that flatter organizations will tend to induce less differential treatment, regardless of whether that is itself the force behind their flattening. In this sense the promotion of fairness and equity in organizations may may coincide with their de-layering and flattening, and these developments might reinforce one another over time.

The monotonicity condition on signal differences in Proposition 2 is plausible in many contexts. Intuitively, this condition means that agents from category x have increasingly higher signals relative to those of category x', given the same benefit index r, i.e. conditional on having the same expected gains in surplus from different tiers. This assumption can be satisfied in settings where agents from category x are relatively over-represented at higher levels of the signal s, e.g. if signals are biased predictors of types for some categories of agents—so they either under-estimate or over-estimate agents' types, when not accounting for the categories—or in settings where some categories of agents have more dispersed conditional signal distributions. This monotone signal differences condition also allows for signal differences to be heterogeneous across benefit indices. For example, it is possible that at lower benefit indices agents of category x' have higher signals than those of category x, while the opposite is true at higher benefit indices; in this case $s(x,r) - s(x',r)$ is negative for r below some threshold and positive for r above that threshold. Example 2 describes a very simple example where this is the case.

Example 2. *Suppose there are 2 tiers, $P = (p_1, p_2)$, with capacities $K = (k_1, k_2)$. The principal wants to assign n agents, n_1 of them with characteristics x^1 and n_2 of them with x^2, with $n_1 + n_2 = k_1 + k_2$.*

Agents' types are distributed normally,

$$t_i^j \sim N(0,1),$$

iid across agents $i \in N$ and characteristics $j = 1, 2$. Their signals are

$$\begin{aligned} s_i^1 &= t_i + \varepsilon_i \quad \text{for category } x^1, \\ s_i^2 &= \alpha(t_i + \varepsilon_i) \quad \text{for category } x^2, \end{aligned}$$

where $\varepsilon_i \sim N(0,1)$, iid for all $i \in N$, and $\alpha > 1$. An agent of type t in position p generates surplus according to some function $v(t,p)$, and all of the assumptions in section 3.1 hold. Note that categories are comparable according to Definition 1, as there exists a one-to-one equivalence between signal distributions.

In this parametric example one can directly study the properties of $r(x,s)$ and $s(x,r)$. Note that $r(x^1,s) = r(x^2,\alpha s)$ for all s, and $s(x^1,r) = \frac{1}{\alpha}s(x^2,r)$ for all r. Therefore the agents' signal differences are

$$s(x^2,r) - s(x^1,r) = (\alpha - 1)s(x^1,r).$$

here signal differences are increasing, since $s(x,r)$ is strictly increasing in r and $\alpha > 1$. That is, as the benefit index increases, category x^2 agents' signals are increasingly biased upwards.

Moreover, for all $r > r(x^1,0)$, $s(x^2,r) - s(x^1,r) > 0$, and for all $r < r(x^1,0)$, $s(x^2,r) - s(x^1,r) < 0$. That is, at higher levels of the benefit index, category x^2 has more favorable signals, while at lower levels of the benefit index category x^1 has more favorable signals.

Following Proposition 1, one can order the agents by their benefit indices, $\{r(x_i, s_i)\}_{i \in N}$*, and then optimally assign them to positions assortatively with respect to their indices. Denote by* i *the agent with the* i^{th} *highest index,* $r_i := r(x_i, s_i)$*, so that the marginal agent assigned to* p_2 *has index* r_{k_2}*.*

Since this example satisfies monotone signal differences, Proposition 2 means that there exists some reference tier $M \in \{1, 2\}$*, such that one can re-allocate capacities upwards or downwards, in a way that strictly reduces all three measures of differential treatment.*

In particular, if $r_{k_2+1} > r(x^1, 0)$*, then the signal differences* $s(x^2, r_{k_2}) - s(x^1, r_{k_2})$ *are positive on the margin, and the reference tier is* $M = 2$*. Therefore* DT_1, DT_2 *and* DT_3 *can all be reduced by moving* k_Δ *positions from* p_1 *to* p_2*, for any* k_Δ *s.t.* $r_{k_2+k_\Delta} > r(x^1, 0)$*. That is, making the organizational structure more "top-heavy" reduces the amount of differential treatment.*

Analogously, if $r_{k_2-1} < r(x^1, 0)$*, then the signal differences are negative on the margin, and the reference tier is* $M = 1$*. Hence* DT_1, DT_2 *and* DT_3 *can all be reduced by moving* k_Δ *positions from* p_2 *to* p_1*, for any* k_Δ *s.t.* $r_{k_2-k_\Delta} < r(x^1, 0)$*. That is, making the organizational structure more "bottom-heavy" reduces the amount of differential treatment.*[13]

Since Example 2 has only 2 tiers, organizational restructuring can reduce differential treatment by re-allocating capacities either upwards or downwards, but not both. More generally, with more than 2 tiers, Proposition 2 means that both of these types of structural changes can co-exist and reduce differential treatment, when the reference tier M is some intermediate tier. Intuitively, for any tier such that signal differences are positive on the margin, re-allocating positions upwards reduces differential treatment, and for any tier such that signal differences are negative on the margin, re-allocating positions downwards reduces differential treatment.

The next proposition adds some additional structure on the agents' characteristics and yields a more specific result in terms of the exact type of organizational differences that reduce differential treatment.

Proposition 3. *Suppose categories can be ordered, s.t. for* $x > x'$*,* $F(t|x, s) \succ_{FOSD} F(t|x', s)$ $\forall\, s$*, and they have monotone signal differences, i.e.* $s(x, r) - s(x', r)$ *is monotone in* r*.*

[13] For the remaining special case where $r_{k_2} > r(x^1, 0)$ and $r_{k_2+1} < r(x^1, 0)$, it is still the case that one can re-allocate capacities relative to some reference tier $M = 1, 2$, in order to reduce DT_1, DT_2 and DT_3, but there are more cases that need to be considered to determine the direction of capacity re-allocation.

If $(P;K)$ is more top-heavy than $(P';K')$, then $(P;K)$ displays less differential treatment than $(P';K')$ according to all metrics, DT_1, DT_2 and DT_3.

The assumption of monotone signal differences by itself means that there will be some tier M such that organizational structures that are flatter around M will yield less differential treatment. The additional assumption of an ordering over categories, which is based on first-order stochastic dominance of the type distributions, implies that the pivotal reference tier must be the lowest one, $M = 1$. Hence top-heavier organizations will generally display less differential treatment.

An ordering over the different categories of agents means that higher category agents have better type distributions, in the sense of FOSD. Such an order may be inherent in some settings, where the agents' characteristics include some factors that directly influence their signals, or it may simply be an empirical regularity that is established when analyzing the relationship between types and signals. For example, in school admissions it may be the case that students with higher household income have better access to test preparation, which allows them to achieve higher admission test scores, conditional on having the same ability as a lower-income student. In this case the household income characteristic provides a natural way to order groups of students, with lower-income students having higher type distributions, conditional on achieving the same admission test score. In such a scenario, a "top-heavier" school district, which has relatively larger school capacities among its most selective schools, would lead to less differential treatment across students with different household incomes.

5 Discussion

This is the first paper to provide an economic theory of how differential treatment varies across organizational structures. I show that under plausible conditions on the agents' signal distributions, flatter, more differentiated, and more top-heavy organizations tend to display less differential treatment. Hence organizations can directly affect the extent of economic inequality among agents with different characteristics, by re-allocating positions across tiers to change their organizational hierarchy, in addition to using preferential treatment policies which indirectly attempt to reduce inequality.

In school districts, for instance, expanding the set of seats available at top or selective schools, thus making the district more top-heavy, can systematically reduce socio-economic inequality in access. This provides an additional policy instrument, which does not directly affect the school district's assignment policy, as is the case with affirmative action policies and reserve and quota systems. Similarly, flattening or de-layering a firm's hierarchy can systematically reduce the gender gap in employment, without explicitly affecting the firm's hiring and promotions policies; i.e. without requiring gender-based policies. This is not to suggest that policies targeting organizational structure should be an alternative to interventions that change the organization's assignment policy. These two different types of tools can operate simultaneously to reduce inequality. In fact, they may be complementary in a sense: if an organization adopts a structure which reduces the extent of differential treatment, then it may require a less dramatic intervention in its assignment policy in order to implement some particular allocation of positions. E.g., a school district with a better-designed structure may need a "smaller" system of reserves and quotas in order to achieve the same level of socio-economic representation within schools.

References

Atila Abdulkadiroğlu. College admissions with affirmative action. *International Journal of Game Theory*, 33(4):535–549, 2005.

Atila Abdulkadiroğlu and Tayfun Sŏnmez. School choice: A mechanism design approach. *American Economic Review*, 93(3):729–747, June 2003.

Dennis J. Aigner and Glen G. Cain. Statistical theories of discrimination in labor markets. *ILR Review*, 30(2):175–187, 1977.

Kenneth J. Arrow. *The Theory of Discrimination*, pages 3–33. Princeton University Press, 1973.

David Bjerk. Glass ceilings or sticky floors? statistical discrimination in a dynamic model of hiring and promotion. *The Economic Journal*, 118(530):961–982, 2008.

Nicholas Bloom and John Van Reenen. Measuring and explaining management practices across firms and countries. *The Quarterly Journal of Economics*, 122(4):1351–1408, 11 2007.

Oguzhan Celebi and Joel P Flynn. Priority design in centralized matching markets. *Available at SSRN 3382580*, 2020.

Stephen Coate and Glenn C Loury. Will affirmative-action policies eliminate negative stereotypes? *American Economic Review*, pages 1220–1240, 1993.

Massimo Colombo and Marco Delmastro. *The economics of organizational design: Theoretical insights and empirical evidence*. Springer, 2008.

Umut Dur and Martin Van der Linden. Capacity design in school choice. *Available at SSRN*

3898719, 2021.

Umut Dur, Scott Duke Kominers, Parag A Pathak, and Tayfun Sönmez. Reserve design: Unintended consequences and the demise of boston's walk zones. *Journal of Political Economy*, 126(6):2457–2479, 2018.

Lars Ehlers and Aytek Erdil. Efficient assignment respecting priorities. *Journal of Economic Theory*, 145(3):1269–1282, 2010.

Lars Ehlers, Isa E. Hafalir, M. Bumin Yenmez, and Muhammed A. Yildirim. School choice with controlled choice constraints: Hard bounds versus soft bounds. *Journal of Economic Theory*, 153:648 – 683, 2014.

Haluk I Ergin. Efficient resource allocation on the basis of priorities. *Econometrica*, 70(6): 2489–2497, 2002.

Hanming Fang and Andrea Moro. Theories of statistical discrimination and affirmative action: A survey. In *Handbook of Social Economics*, volume 1, pages 133–200. Elsevier, 2011.

Roland G Fryer. Belief flipping in a dynamic model of statistical discrimination. *Journal of Public Economics*, 91(5-6):1151–1166, 2007.

Maria Guadalupe, Hongyi Li, and Julie Wulf. Who lives in the c-suite? organizational structure and the division of labor in top management. *Management Science*, 60(4):824–844, 2014.

Isa E Hafalir, M Bumin Yenmez, and Muhammed A Yildirim. Effective affirmative action in school choice. *Theoretical Economics*, 8(2):325–363, 2013.

Shelly J Lundberg and Richard Startz. Private discrimination and social intervention in competitive labor market. *American Economic Review*, 73(3):340–347, 1983.

George J Mailath, Larry Samuelson, and Avner Shaked. Endogenous inequality in integrated labor markets with two-sided search. *American Economic Review*, 90(1):46–72, 2000.

Paul R Milgrom. Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics*, pages 380–391, 1981.

Andrea Moro and Peter Norman. A general equilibrium model of statistical discrimination. *Journal of Economic Theory*, 114(1):1–30, 2004.

Edmund S Phelps. The statistical theory of racism and sexism. *American Economic Review*, pages 659–661, 1972.

Raghuram G Rajan and Julie Wulf. The flattening firm: Evidence from panel data on the changing nature of corporate hierarchies. *The Review of Economics and Statistics*, 88(4): 759–773, 2006.

Raghuram G. Rajan and Luigi Zingales. The Firm as a Dedicated Hierarchy: A Theory of the Origins and Growth of Firms*. *The Quarterly Journal of Economics*, 116(3):805–851, 08 2001.

Emil Temnyalov. An information theory of efficient differential treatment. *American Economic Journal: Microeconomics*, 2021.

Julie Wulf. The flattened firm: Not as advertised. *California Management Review*, 55(1): 5–23, 2012.

6 Appendix: Proofs

To be added.